AD-A055 833 WOODS HOLE OCEANOGRAPHIC INSTITUTION MASS F/G 9/2
A DIGITAL TAPE FORMAT FOR WOODS HOLE OCEANOGRAPHIC INSTITUTION --ETC(U)
MAY 78 R C MILLARD, A BLUMER, N GALBRAITH N00014-76-C-0197
UNCLASSIFIED WHOI-78-43 NL

| OF |
AD
A055833

END
DATE
FILMED
8 -78
DDC

UNCLASSIFIED 5/78

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1 REPORT NUMBER: WHOI-78-43

2 GOVT ACCESSION NO.

3 RECIPIENT'S CATALOG NUMBER

4 TITLE (and Subtitle): A DIGITAL TAPE FORMAT FOR WOODS HOLE OCEANOGRAPHIC INSTITUTION CTD DATA.

5 TYPE OF REPORT & PERIOD COVERED: Technical Report

6 PERFORMING ORG. REPORT NUMBER

7 AUTHOR(s): R. C. Millard, A. Blumer, N. Galbraith

8 CONTRACT OR GRANT NUMBER(s): N00014-76-C-0197

9 PERFORMING ORGANIZATION NAME AND ADDRESS
Woods Hole Oceanographic Institution
Woods Hole, MA 02543

10 PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR 083-400

11 CONTROLLING OFFICE NAME AND ADDRESS
NORDA
National Space Technology Laboratory
Bay St. Louis, MS 39529

12 REPORT DATE: May 1978

13 NUMBER OF PAGES: 36 (12) 41 p.

14 MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15 SECURITY CLASS. (of this report): Unclassified

15a DECLASSIFICATION/DOWNGRADING SCHEDULE

16 DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

19 KEY WORDS (Continue on reverse side if necessary and identify by block number)

1. Magnetic tape format
2. CTD Data Storage
3. 16-bit mini-computer

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

A new digital data tape format has been developed to be used for data from the WHOI/Brown CTD microprofiler. CTD-78 Version 1 as detailed in this report is designed as a flexible and expandable internal data format adapted to the Hewlett-Packard 2100/21MX series 16-bit mini-computers currently used to transcribe and process CTD data at sea. The ten record types presently used in this multifile tape format store CTD and associated water sample data and the labeling information needed to convert the data to physical units. Record types are identified by unique keywords.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

78 06 28 116

UNCLASSIFIED 5/78

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

381 004

WHOI-78-43

A DIGITAL TAPE FORMAT FOR WOODS HOLE OCEANOGRAPHIC INSTITUTION CTD DATA

by

R. C. Millard, A. Blumer,
and N. Galbraith

WOODS HOLE OCEANOGRAPHIC INSTITUTION
Woods Hole, Massachusetts 02543

May 1978

TECHNICAL REPORT

Prepared for the Office of Naval Research under Contract N00014-76-C-0197; NR 083-400.

Approved for public release; distribution unlimited.

Approved for Distribution

Valentine Worthington, Chairman
Department of Physical Oceanography

Table of Contents

Abstract

A new digital data tape format has been developed to be used for data from the WHOI/Brown CTD microprofiler. CTD-78 Version 1 as detailed in this report is designed as a flexible and expandable internal data format adapted to the Hewlett-Packard 2100/21MX series 16-bit mini-computers currently used to transcribe and process CTD data at sea. The ten record types presently used in this multifile tape format store CTD and associated water sample data and the labeling information needed to convert the data to physical units. Record types are identified by unique keywords.

Introduction

The following is a description of a digital tape format developed for the WHOI/Brown CTD microprofiler. This report details an internal tape format called CTD-78 adapted to the Hewlett-Packard 2100/21MX series 16-bit mini-computers currently used at Woods Hole Oceanographic Institution (WHOI) to transcribe and process CTD data at sea. The tape format structure is designed to be expandable, as new uses and requirements are identified. Further record types beyond the ten described in this report can be added and the tape format can accommodate data channels added to the instrumentation or extended data resolution without tape format modification.

The information contained in the station label records is an extension of both the Buoy Tape Format (Maltais, 1969) and GATE formats (GATE Report 13, 1974), since versions of WHOI CTD data are transcribed into both of these tape formats. The structure of the CTD-78 tape format is adapted from the Buoy Tape Format.

The report is divided into four sections. First a description is given of the tape structure and file contents. The second section gives the purpose and a description of the information content of each currently defined record. Third is a section describing the specially defined fields. The fourth section contains appendices; data conversion algorithms, glossary of terms, and block diagrams.

I. Tape Structure

CTD-78 is a multifile tape format, with each record identified by a keyword. The first file of every tape contains the Tape Header record or records if historic tape headers are retained. After this file are the station files, one file per CTD station or cast. Files are separated by EOF marks, and there is a double EOF after the last station file on the tape.

Within each CTD station file, there are two kinds of records: labeling records which document the station file and data records containing CTD or water sample data. The first record of each station file is a Station File Header record, and the last is always a File Trailer record. Any condition other than this is an error condition, probably indicating an incomplete file. Between the station file header and file trailer records are records in the following order: water sample scale factor records; water sample data records; CTD scale factor records; CTD data records. The water sample scale factor and data records are added when this data is available. A complete set of Scale Factor records must always precede the Data records they are describing, and there can be at most two sets in a file, one describing ordinary Data records, and the other for Water Sample Data records.

Comment records can be interspersed at will throughout the file, except that they may not occur before a Station Header or after a File Trailer. Historic label records are carried in the station file to document changes made to label records. Historic records must precede CTD data records. Usually they will be adjacent to the label record type they are documenting.

An example of the tape structure is shown in Figure 1. The example is for a data set which has been processed through editing and derived data (computation of salinity and oxygen) steps. The station label records represent only .5 percent of the total file length[1] for a typical 1000 data record lowering to 5000 meters.

---

[1]The total file length for a station with 1000 data records is about 279 feet at a recording density of 800 BPI. See Table 1 for the physical record length of various record types.

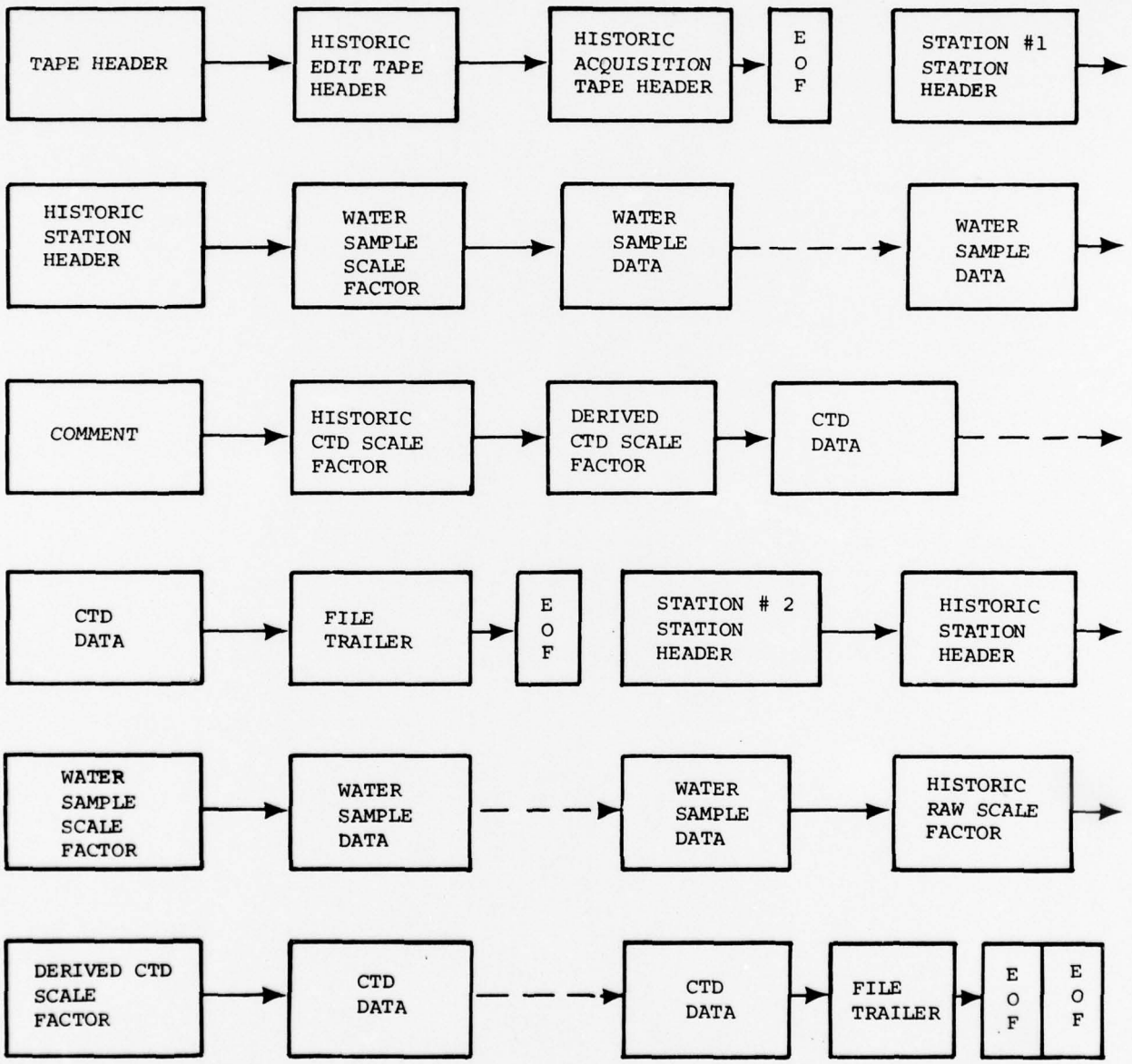


Figure 1. Sample tape layout.

II. Record Types

There are ten records defined in the CTD-78 format Version 1 as listed in Table 1. Different record types are distinguished by keywords, which are always the first word of the record, and are two's complement integers. CTD Data records (but not Water Sample Data records) have consecutively increasing positive integers for the keyword; this also serves as the record number. All other records are distinguished by zero and negative keywords. The non-positive keywords are divided into current record types and historic record types. A historic record keyword is formed by subtracting 256 from the current record type keyword (CTD Data records are not allowed historic records). Version 1 has 10 record types listed in Table 1.

Labeling records contain three distinct categories of information: 1) Station and processing documentation, 2) Data identifiers and conversion constants, 3) Program control words. The program control words allow processing programs to manipulate the expandable parts of the tape format such as data scan length, data descriptor length, increased data resolution, and the different record types (keywords).

Data and Scale Factor records are 1032 HP words long; all other records are 90 HP words long. All the 90-word records have the last 36 words (72 bytes) reserved as a comment field. The comments will differ in nature from record type to record type, but they will always be 72 bytes of ASCII characters, padded at the end with blanks.

Information fields have the following default values: ASCII fields are blank; most two's complement integers (TCI) and floating point are zero; and unsigned integer (USI) have all bits set to one. Unused fields are set equal to two's complement integer zeroes. Special default values are set for certain fields. Time,latitude, longitude, wind speed, minimum pressure and timer unit words are all -9999.

This section gives the purpose of each currently defined record type followed by a brief description of the information content. Those items which require more detailed explanation have an asterisk appended and a complete definition in Section III.

Table I. Record Types

Keyword	Record length bytes/H.P. words inches at 800 bpi*	Record Name	Description
+N	2064/1032	CTD data	Data digitized by CTD instruments is stored. N is the consecutive record number
0	180/90 .225	Tape header	Always the first record on tape. It identifies the tape
-1	180/90	File trailer	Always the last record of a station file
-2	180/90 .225	Station file header for acquisition	Always the first record of a station file. Contains station cataloging information.
-3		Station file header for edit	
-4		Raw data scale factor	Always immediately precedes associated data records
-5	2064/1032	Derived data scale factor	
-6	2.58	Water sample scale factor	Contains conversions to physical units
-7	2064/1032	Water sample data	Data collected by water sampling device with associated CTD data
-8	180/90	Comments	Plain English text - extension of comments found in other records
-9 through -255	180/90 or 2064/1032	'Open'	These keywords are available for record types to be defined in the future
-256 through -32768	180/90 2064/1032	Historic	Keyword formed by subtracting 256 from above keywords. Allows previous versions of records to be carried along to document changes to information

*Note - Interrecord gaps are nominally .75 inches long and End of File marks are nominally 3 inches long.

A. Tape Header Record

Each tape begins with a tape header record which contains the tape name, date created, project code, tape format version number, and a comment relevant to the overall objectives of the tape. The tape header record or records comprise the first file on a tape followed by an end-of-file mark.

A tape format version number is included and this report documents tape format version number one. The unused fields in Version 1 records provide some room for expansion in future versions, beyond which new record types may be created.

Tape Header Record

Word #	Item	Description	Type	Details Page #
1	KEYWORD	Equals 0 for the tape header record	TCI	
2	PROJECT CODE	This number uniquely identifies stations of common purpose	TCI	
*3-5	DATE TAPE CREATED	The year, month, day that this tape was created	TCI	22
6-7	TAPE NAME	A unique 4-character name given the tape	ASCII	
8-9	SOURCE TAPE NAME	The 4-character input tape name. First input tape name for merged files	ASCII	
10	FORMAT VERSION	A positive integer to uniquely link the format version to documentation	TCI	
11-54	UNUSED	Unused in Format Version 1	TCI	
55-90	COMMENT	Comments on the objective of the tape	ASCII	

*See Section III for specially defined fields.

B. Comment Record

Although comments are permitted in three record types (tape header, station header, and file trailer), another record type devoted only to plain English or native language comments is included to extend the area available for comments. The 72 ASCII characters assigned to comments occur in the same position in each of the four record types.

Comment Record

Word #	Item	Description	Type
1	KEYWORD	Equals -8 for the Comment record	TCI
2-54	UNUSED	Unused in Format Version 1	TCI
55-90	COMMENT	A plain English comment	ASCII

C. Station File Header Record

This record type contains the bookkeeping and processing information to identify the station file, such as ship, cruise, station number, date, time, position, and data version. Processing programs can use the information stored in this record to sort and catalog data files.

There are acquisition and edited station file header records distinguished by keywords -2 and -3 respectively. The acquisition station file header uses only some of the information fields (words 1-18, 28, 37, and the comment), since the remaining information is not available at the start of the station.

Every acquisition station file within a cruise is identified by a unique station number. Should the acquisition station file be divided into smaller files during editing (for example, repeated down-up casts as separate files), then each file is uniquely identified by a combination ship, cruise, station, cast number, and data version.

Station File Header Record

Word #	Item	Description	Type	Details Page #
1	KEYWORD	Equals -2 for the acquisition file header Equals -3 for the edited file header	TCI	
2	PROJECT CODE	A unique number to identify stations of common purpose	TCI	
*3	SHIP CODE	A 2-character mnemonic to uniquely identify the ship	ASCII	24
4	CRUISE NUMBER	A unique cruise or expedition number	TCI	
5	STATION NUMBER	A unique station number assigned consecutively within the cruise	TCI	
*6	DATA VERSION	A code that indicates the stage of data processing	TCI	20
*7-9	DATE OF STATION (START)	The year, month, and day that the station was collected	TCI	22
*10	TIME (START)	The Greenwich Mean Time at the start of the station	TCI	23
*11-12	LATITUDE (START)	The beginning latitude of the station	TCI	23
*13-14	LONGITUDE (START)	The beginning longitude of the station	TCI	23
15	WORDS PER SCAN	The number of HP words per scan in the CTD data records	TCI	
16	SCAN RATE	The number of CTD data scans per sec (100 × number of scans/sec)	TCI	
*17	TIME UNIT FREQUENCY	The number of computer timer pulses per second multiplied by 100	TCI	23
18	PRESSURE SAMPLING INTERVAL	The sampling interval in uniform pressure series; 10 times the number of decibars	TCI	

*See Section III for specially defined fields.

Station File Header Record (Contd)

Word #	Item	Description	Type	Details Page #
*19-20	LATITUDE (END)	The latitude at the end of the station	TCI	23
*21-22	LONGITUDE (END)	The longitude at the end of the station	TCI	23
*23	TIME (END)	The Greenwich Mean Time at the end of the station	TCI	23
24	MINIMUM PRESSURE	The minimum pressure recorded in the file (nearest decibar)	TCI	
25	MAXIMUM PRESSURE	The maximum pressure recorded in the file (nearest decibar)	TCI	
*26-27	JULIAN DAY	The Julian day number has an offset of 244×10^4	TCI	22
28	INSTRUMENT	A unique identification number for the CTD underwater unit	TCI	
*29	QUALITY FLAG	Indicator for the overall data quality of the file	TCI	19
*30-32	EDIT DATE	The year, month, day of the station file edit	TCI	22
33	WATER SAMPLES	The number of water samples collected with this station	TCI	
*34	POSITION METHOD	A 2-character mnemonic indicating the method used to determine station position	ASCII	24
35	WIND SPEED	Anemometer reading, in meters/sec	TCI	
36	WATER DEPTH	The corrected bottom depth to the nearest meter	TCI	
*37	STATION TYPE	A 2-character mnemonic describing the collection method	ASCII	25
38	CAST NUMBER	A number assigned consecutively within each station	TCI	

*See Section III for specially defined fields.

Station File Header Record (Contd)

Word #	Item	Description	Type
39-54	UNUSED	Unused in Version 1	TCI
55-90	COMMENT	The first 10 characters are for the last name of person creating the file. The last 8 characters are reserved for program version	ASCII

D. Scale Factor Record

The scale factor records contain the parameters necessary to convert the data into physical units. The keywords distinguish three types of scale factor records: 1) Raw data refers to data as digitized by the instrument; 2) Derived data has conversion algorithms applied to the measured parameters to form variables such as salinity and oxygen; 3) Water sample data allows hydrographic data to be stored along with the CTD data.

The scale factor records are 1032 words long and consist of program control information plus sets of variable descriptors, each describing an HP data word.

Words 1 through 8 of the scale factor record are program control words. Word 1 is the keyword equal to -4, -5, or -6. Word 2 gives the number of variable descriptors in the record. The length of the variable descriptor in HP words is contained in word 3 and the number of floating point words of the variable descriptor in word 5. Word 4 gives the number in HP words in a data scan. Words 6 through 8 are currently unused.

Scale Factor Record

(Program Control Words = Words 1-8)

Word #	Item	Description	Type
1	KEYWORD	Equals -4 for instrument data Equals -5 for derived data Equals -6 for water sample data	TCI
2	NUMBER OF VARIABLES	The number of variable descriptors in the record	TCI
3	DESCRIPTOR LENGTH	Length of the variable descriptor in HP words. Equals 34 for format Version 1	TCI
4	WORDS PER SCAN	The number of HP words in a data scan	TCI
5	NUMBER OF FLOATING POINT VALUES	The number of FP values in the descriptors. Equals 5 for Version 1	TCI
6-8	UNUSED	Unused in Version 1	TCI
9-1032	VARIABLE DESCRIPTORS	Variable descriptor 1: applies to first HP word of data scan Variable descriptor 2: applies to second HP word of data scan . . . Variable descriptor N: applies to N^{th} HP word of data scan	

The variable descriptors begin at word 9. Each variable descriptor is associated with an HP word within a data scan in the same order. Each variable descriptor is an even number of HP words in length to allow storing of floating point information. The variable descriptor is expandable in length to allow addition of information in later tape format versions. New information is added to the middle of the variable

descriptors with floating point information added adjacent to the presently defined floating point words. The end of the record after the last variable descriptor is padded with two's complement zeroes.

N^{th} Variable Descriptor in Scale Factor Record

To find the starting HP word number of this descriptor, subtract 1 from the variable number, then multiply by the descriptor length (scale factor record word 3) and add 9.

Relative Word	Item	Description	Type	Details Page #
1-4	VARIABLE NAME	The name of the variable, in English, 8 characters	ASCII	
5-9	DIMENSIONAL UNITS	The units of variable, in English, 10 characters	ASCII	
*10	VARIABLE IDENTIFIER	A 2-character mnemonic which identifies the variable	ASCII	25
11	LAG CORRECTION WINDOW	The number of scans to use for sensor lag correction	TCI	
*12	QUALITY FLAG	An indicator of the quality of the sensor calibration	TCI	19
*13	BITS RESOLUTION	The number of bits resolution for instrument variables	TCI	21
14	DELTA EDIT CRITERION	The edit criterion for differencing, in digitizer units	TCI	
15	SENSOR NUMBER	The identification number of the sensor	TCI	
*16-18	DATE OF CALIBRATION	The year, month, day of calibration of the sensor	TCI	22
19	SIGN BIT WORD	The location within the data scan of the sign bit word for this variable	TCI	
■20	SIGN BIT MASK	The mask applied to the sign bit word to extract the sign bit: the bit is on for negative data	TCI	

*See Section III for specially defined fields.

■See Appendix I for Hewlett-Packard CTD data manipulating routines.

N^{th} Variable Descriptor in Scale Factor Record (Contd)

Relative Word	Item	Description	Type
▪ 21	LEAST SIGNIFICANT BITS WORD	The location within the data scan of least significant bits for this variable	TCI
▪ 22	LEAST SIGNIFICANT BITS MASK	The mask applied to L.S.B. word to extract the least significant bits	TCI
23	DIGITIZING PERIOD	The period of updating the variable's data, in scans	TCI
▪ 24	DATA MASK	The mask applied to instrument data before outputting to tape to insure zero for unused high order bits	TCI
FROM END:			
4-5	ATTRIBUTES 1 AND 2	To be assigned for scaling or other data attributes as needed	F.P.
3	SLOPE	The slope to be applied to data	F.P.
2	BIAS	The bias to be applied to data (in physical units)	F.P.
1	SENSOR LAG	The time constant of the sensor, in seconds	F.P.

▪ See Appendix I for Hewlett-Packard CTD data manipulating routines.

E. CTD Data Records

The CTD data records are 1032 HP words long, which makes them compatible with early CTD formats (Tollios et al., 1971). The first 8 words pertain to the record as a whole. The rest of the record is made up of data scans of unsigned integer words (data/program generated), with each scan having the same number of words, and always in the same order as

the data descriptors in the scale factor record. The number of words in a data scan is specified both in word 15 of the File Header record and in word 4 of the CTD scale factor record. Four kinds of unsigned integer words are stored in a data scan: 1) most significant bits of a variable; 2) sign bits, 3) least significant bits of a variable; and 4) program generated words (quality/time). Instrument data is written in the data record as telemetered with the following restrictions. The signs of the data are stored in a separate word. The least significant bits are always consecutive bits within an unsigned integer word with the least significant data bit in the least significant part of the unsigned integer word. When CTD data is logged as a time series, the data scans within each tape record are written uniformly in time. The time base within data records is determined by the rate the instrument sends data. A time word generated by a clock in the data logging computer determines the time information between data records which is stored in words 2 and 3 of each data record. Time discontinuities may exist between data records. When CTD data is converted to a uniform pressure series a data scan is associated with each pressure interval. The identification of gaps in the data is discussed in the special definitions section under the data quality word.

Algorithms for decoding the unsigned integer, sign and extended data resolution bits are given in Appendix I with an example of their use. Remaining words after the last complete scan in a data record are padded with all bits set to one (65535).

CTD Data Records

Word #	Item	Description	Type	Details Page #
1	KEYWORD	A positive integer, equals the CTD data record number	TCI	
*2	TIME	The Greenwich Mean Time at the start of the data record	TCI	23
*3	TIMER UNITS	The number of timer units since the last minute of word 2 above	TCI	23
4	STATION NUMBER	A unique Station Number assigned consecutively within a cruise	TCI	
5	RECORD TAG	An acquisition stage event marker. If non-zero, flags records as having associated water sample data, for example	TCI	
6	NUMBER OF ERRORS	The number of errors detected in this record	TCI	
7	NUMBER OF SCANS	The number of data scans in this record	TCI	
8	DATA CHECKSUM	Checksum on the data in the rest of the record	TCI	
		$CHKSUM = \sum_{N=9}^{1032} IDATA(N)$	TCI	
9-1032	CTD DATA	The data is organized in scans, one after the other, until the scan number specified by word 7, after which the record is padded with USI 65535	USI	

*See Section III for specially defined fields.

Typical Raw Data Scan

Pressure	Temperature	Conductivity	Sign bits	Oxygen Current	Oxygen Temperature	Quality Raw

Typical Derived Data Scan

Pressure	Temperature	Salinity	Sign bits	Oxygen	Quality Raw	Quality Salinity

F. Water Sample Data Record

The water sample data record stores data collected from water sampling devices (e.g., Rosette samplers) used with the CTD. Since certain water sample data is used to check the CTD calibrations, this record usually contains associated CTD data stored in each scan of the water sample data record. The water sample data starts at word 9. Missing observations within a scan are coded as 65535 as are remaining words after the last data scan.

Water Sample Data Records

Word #	Item	Description	Type
1	KEYWORD	Equals -7 for Water Sample Data record	TCI
2-6	UNUSED	Unused in Version 1	TCI
7	NUMBER DATA SCANS	The number of data scans in this record	TCI
8	DATA CHECKSUM	Checksum on the data in the rest of the record $CHKSUM = \sum_{N=9}^{1032} IDATA(N)$	TCI
9-1032	WATER SAMPLE DATA	The data is organized in scans, the record padded with USI 65535	USI

Water Sample Data records look exactly the same as CTD Data, except that the keyword is always -7, and words 2-6 are unused.

Typical Water Sample Data Scan

P_{CTD}	T_{CTD}	C_{CTD}	Sign bits CTD	OC_{CTD}	OT_{CTD}	Salinity W.S.*	Oxygen W.S.*	Silicate W.S.*	Quality

*W.S. = water sample.

G. File Trailer Record

The station file trailer record stores information available at the completion of a station file. The file trailer record is always the last record of a station file and is used as a data terminator by the processing programs.

File Trailer Record

Word #	Item	Description	Type	Details Page #
1	KEYWORD	Equals -1 for File Trailer Record	TCI	
*2	TIME (END)	GMT on a 24-hr clock, at the end of the station	TCI	23
*3	TIMER UNITS	The number of timer units elapsed since the last minute in word 2	TCI	23
4	ABORT FLAG	Equals 0 if the station terminated normally Equals 1 if the station is aborted	TCI	
5	SYNC ERRORS	The number of scans with synchronization errors in the station file	TCI	
6	EDIT ERRORS	The number of scans with edit error flags	TCI	
*7	QUALITY FLAG	An indicator of the overall quality of the file	TCI	19
*8-9	LATITUDE (END)	The latitude at the end of the station	TCI	23
*10-11	LONGITUDE (END)	The longitude at the end of the station	TCI	23
*12-14	DATE OF STATION (END)	Year, month, day at the end of the station	TCI	22
15-54	UNUSED	Unused in Version 1	TCI	
55-90	COMMENTS	Final comments on file	ASCII	

*See Section III for specially defined fields.

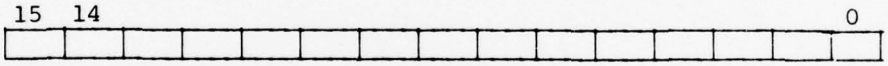
III. Specially Defined Fields

Certain words briefly described in the various record types require further definitions in terms of the information to be coded.

A. Data Quality Words

Data quality words appear in several record types. All quality codes are defined such that a zero value means good; a positive value indicates a problem, and a negative value indicates a corrected problem. Comments can be used to clarify error conditions.

Station file header word 29 indicates the overall quality of the data file combining both spurious data and calibration problems. The bits in this word refer to the first fourteen variables within a data scan where off condition means good and on means questionable. Bit 14 refers to variable 15 and greater of data scan. A negative value of the quality word indicates errors that the Editor can correct have been corrected.



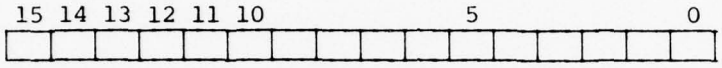
bit	On Condition Indicates:
15	errors corrected
14	variable 15 or beyond has questionable data
.	
.	
.	
0	variable 1 has questionable data

Scale factor variable descriptors word 12 indicates the quality of the sensor calibration. A zero means good and 1 means questionable.

File Trailer word 7 indicates the variables contributing to the error content of the data file using the same flagging scheme of the station header quality word.

Data Scan Quality Word. The scan quality word occurs with each scan of CTD data. The bits within the word are used to indicate variables failing the edit criteria according to the following scheme. For more than 13 variables a second quality word is added with the least significant bit flagging variable 14, etc.

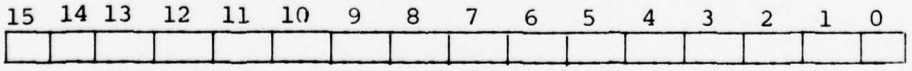
Quality word description



Bit:	On Condition Indicates:
15	Spurious data corrected
14	Frame synchronization error detected for time series data; gap in data for uniform pressure series
13	Invalid edit criteria - also sets bit 15 on
12	Error in 13th variable of scan
.	
.	
.	
0	Error in first variable of the scan

B. Data Version

Station header word 6 - The data version is used to keep a record of the processing steps of a station file. Groups of bits are used to store a count of the number of edits.



Version	Bit(s)	Code
Acquisition	0	0 = direct from cable 1 = re-read audio tape
	1-2	Unused
Raw Data Edit	3-5	Number of passes
Derived Data Edit	6-8	Number of passes
Raw Data Error Correction	9	1 = errors corrected
Derived Data Error Correction	10	1 = errors corrected
Raw/Derived Data Flag	11	0 = Raw data 1 = Derived data
Uniform Pressure/Time	12	0 = Uniform time 1 = Uniform pressure
Average	13	1 = Averaged data
Copy	14	1 = Copied data
Data	15	**0 = Preliminary** **1 = Final**

For example, an octal data version of 37121 represents a digital transcription from audio tape of a data file in which the raw data was edited twice and salinity derived and edited once. Both raw and salinity errors were corrected and data is an averaged uniform pressure series.

C. Bits Resolution

Scale factor record variable descriptor word 13 - The number of bits resolution of a sensor variable. The meaning of this word depends upon the type of variable being described. The magnitude can be tested by processing programs to determine how to handle the variable. For example, only data variables with a positive value in this field should be corrected for sensor lags.

1. Instrument variables with a sensor variable identifier (PR, TE,...). Contains the number of bits resolution of the sensor digitization including sign bit.

2. Instrument variables without a sensor identifier (sign and least significant bits). Contains the number of bits ($\leq$16) of the H.P. word used. Stored as a negative number.

3. Program generated variables (quality and time). Always zero.

D. Date

Calendar date - tape header words 3-5; station file header words 7-9 and 30-32; variable descriptor words 16-18; and file trailer words 12-14.

The date is expressed as three TCIs, the first of which is the last two decimal digits of the year, the second is the month number within the year, and the third is the day number within the month.

Examples:

Date	Word 1	Word 2	Word 3
January 1, 1976	76	1	1
September 12, 1968	68	9	12

Julian Day Number-Station Header Words 26-27

The Julian Day is formed by combining words 26 and 27. Word 26 is the Julian day number offset when multiplied by 10^4. The Julian Day number offset is 2440000. The variable part of the Julian day number is stored in word 27. Algorithms for converting word 27 to Gregorian date and back to Julian day number are given in Appendix I. The algorithms are coded for the H.P. series 2100/21MX computer after Tantzen (1963) and are valid from March 1, 1968 thru December 31, 1999.

E. Time

Station header words 10, 17, and 23: Data records and file trailer words 2 and 3

1. When resolution is only needed to minutes, the Greenwich Mean Time (GMT) is stored as a Two's Complement Integer as follows: the hour number on a 24-hour clock is multiplied by 100 and added to the number of minutes within that hour. For example, 16 minutes past 2 o'clock in the afternoon, GMT, would be represented as 1416. This form is easy to print using decimal format conversion in FORTRAN (such as F5.2), but more difficult to work with for computing elapsed times.

2. When resolution within minutes is needed, it is given in terms of the number of timer pulses elapsed since the last minute. Word 17 of the File Header record gives 100 times the number of pulses per second, so the following gives the number of seconds elapsed since the last minute: (100*INCR)/W17 where INCR is the number of pulses elapsed since the last minute, and W17 is the number from word 17 of the File Header record.

3. The time of observation of an individual scan within a data record requires the scan rate from word 16 of the file header record to compute. The time in seconds, from the beginning of the record, of the data scan in position ISCAN within the record is: (100*ISCAN)/W16 where W16 is the number from word 16 of the file header record.

F. Position

Latitude and Longitude - Station header words 11-14 and 19-22, file trailer words 8-11

Latitudes north of the equator are positive, south are negative, and longitudes east of Greenwich are positive, west are negative. Latitude and longitude are each expressed as a pair of two's complement integers which are either both positive, both negative, or one of them is zero. The first word is the number of degrees, and the second is the number of minutes multiplied by 100.

Examples:

Position	Word 1	Word 2
39 degrees, 27.75 minutes north of the equator	39	2575
40 degrees, 17.53 minutes south of the equator	-45	-1753
177 degrees, 5.45 minutes east of the Greenwich Meridian	177	545
70 degrees, 7 minutes west of the Greenwich Meridian	-70	-700

G. Mnemonics

Two character mnemonics are used where it is desirable to code a single piece of information that can be both printed and tested in an integer form. The rules for creating a mnemonic are to check previously created mnemonics, then form a unique mnemonic using the first 2 characters of a single word or the first character of two words.

1. Ship Code Mnemonics - Station header word 3

Chain - CH

Atlantis II - AT

Knorr - KN

Oceanus - OC

Kurchatov - KU

Vernadsky - VE

Thomas Thompson - TT

2. Positioning Method - Station header word 34

Unknown - UN

Dead Reckon - DR

Satellite - SA

Loran A - LA

Loran C - LC

Omega - OM

3. Station Type - Station header word 37

Cable lowered (down and up) - CL
Cable down - CD
Cable up - CU
Ship towed - ST
Free fall - FF

4. Variable I.D. - Each scale factor descriptor word 10

Pressure - Pr
Temperature - TE
Conductivity - CO
Timer units - TU
Oxygen current - OC
Oxygen temperature - OT
Salinity - SA
Oxygen - OX
Nephilometer - NE
Least significant bits - LS
Sign word - SW
Quality - QU
Quality one - Q1
Quality two - Q2

Restrictions

An expandable tape format for storing CTD data on Hewlett-Packard mini-computers has been described. The following limitations exist to the growth of the tape format:

1. The maximum length of a data scan is 1024 H.P. words.
2. The maximum data resolution is limited by the H.P. floating point to 23 bits and the sign.
3. The maximum number of defined label records is 256.
4. The maximum number of historic label records of each type is 128.
5. The maximum number of data records is 32767.
6. The maximum variable descriptor length is 1024 H.P. words.

Appendix I

Hewlett-Packard CTD Data Manipulating Functions

The WHOI/Brown CTD currently has a 16-bit digital to analog converter (Brown, 1974; Brown and Morrison, 1978). For certain variables the sign of the digitized number is also determined, extending the resolution to 17 bits (Brown and Morrison, 1978). The Hewlett-Packard series 2100/21MX computers require a special routine to convert the 16-bit unsigned integer data to floating point, interpreting the integer sign bit as the most significant bit. Function R makes this conversion. The sign of the digitized number is stored in a separate word of the data scan. The location of this sign word and a mask to locate the appropriate bit containing the sign (1 equals negative) are given in the variable descriptor section of the scale factor record for each variable. Function ISNFP accesses the sign information.

Design has begun on an 18-bit digitizer version of the WHOI/Brown CTD microprofiler. The least significant bits (LSBs) component of each digitization is stored in another word of the data scan. Again the location of the word containing the LSBs and a mask to locate the appropriate bits within the word are contained in the variable descriptor section of the scale factor record. Function FPLSB extracts these bits and scales by the resolution implied by the LSB mask so the value returned is always less than one. For example, given an LSB value of 2 octal and a mask of 3 octal, function FPLSB will return a value of 0.5 since the mask implies a resolution of 1 part in four.

Data Conversions to Physical Units

This example converts the N^{th} variable of a data scan ISCAN(N).

Two indices are required to obtain information from the N^{th} scale factor record variable descriptor.

SFW3 is the length of each variable descriptor (word 3 of the scale factor record).

```
ISFW = 9 + (N-1) * SFW3          integer information index
 SFW = (8 + N * SFW3)/2          real information index
```

The conversion of digitizer data to an H.P. floating point number (RDATA) including sign and least significant bits information using the H.P. data manipulating functions is as follows:

$$RDATA = \overbrace{FLOAT(ISNFP(ISFW+19,ISFW+20))}^{\text{Sign}} * \overbrace{(R(ISCAN(N))}^{\text{USI}} + \overbrace{FPLSB(ISFW+21,ISFW+22))}^{\text{Least Significant Bits}}$$

For a linear conversion of the floating point data to physical units

$$EDATA = (SFW-2)*RDATA + (SFW-1)$$

Function R - Interprets H.P. Two's complement integer sign bit as most significant bit for unsigned integer data. Occupies 37 locations. A real result is returned (Tollios et al., 1971).

```
      FUNCTION R(N)

      1A = 77777B
      IF(N)1,2,2
    1 R = FLOAT(IAND(N,IA)) + 32768
      RETURN
    2 R = FLOAT(N)
      RETURN
      END
```

Function ISNFP - Returns the data sign (±1). Occupies 31 locations. An integer result is returned.

```
      FUNCTION ISNFP(IUW,IMASK)
C   RCM DEC 77
      J = IAND(IUW,IMASK) - 1
      IF(J)1,2
    1 ISNFP = 1
      RETURN
    2 ISNFP = -1
      RETURN
      END
```

Function FPLSB - Returns the least significant bits of the data scaled by the magnitude of the LSB mask plus one. Result always less than one. The technique requires a Two's complement computer. Occupies 58 locations. A real result is returned.

```
      FUNCTION FPLSB(IUW,LMASK)
C  RCM DEC 77
      J = -LMASK
      J = IAND(J,LMASK)
      M = LMASK/J+1
      ISB = IAND(IUW,LMASK)/J
      FPLSB = FLOAT(ISB)/FLOAT(M)
      RETURN
      END
```

H.P. Julian Day Number/Gregorian Date Conversions

```
      FUNCTION KDAY(ID,IM,IYR)

C CONVERT GREGORIAN DATE TO JULIAN DAY
C USES LAST 4 DIGITS OF JULIAN DAY.  ADD 2440000 TO GET THE
C FULL JULIAN DAY.
C NPF
C
C JULY 12 1975
C
      IY = IYR - 68
      IF(2-IM)10,20
   10 M = IM - 3
      GO TO 30
   20 M = IM + 9
      IY = IY - 1
   30 IDY = (1461.*FLOAT(IY))/4.0
      KDAY = IDY + (153*M+2)/5+ID-84
      RETURN
      END

      SUBROUTINE KDATE(KD,ID,M,IY)

C CONVERT JULIAN DAY TO GREGORIAN DATE
C NPF
C
      K = KD+84
      IY = (4.0*FLOAT(K) - 1.0)/1461.0
      ID = 4*K-1-1461*IY
      IY = IY + 68
      ID = (ID+4)/4
      M = (5*ID-3)/153
      ID = 5*ID-3-153*M
      ID = (ID+5)/5
      IF(M-10)20,10
   10 M = M-9
      IY = IY + 1
      RETURN
   20 M = M+3
      RETURN
      END
```

Appendix II

Glossary of Terms

BPI	- Bits per inch - the digital data tape recording density.
Byte	- 8 bits, one ASCII character, or half an HP word.
Data Scan	- A set of HP words (variables, q.v) in an order determined by the variable descriptors (q.v) in the Scale Factor record(s).
Descriptor	- See variable descriptor, below.
EOF	- End of file mark.
HP Word	- A 16-bit field, used here because a word of this length is used by Hewlett-Packard 2100/21MX series computers.
Keyword	- The first word of a tape record, identifying what type of record it is.
Mask	- A word that specifies which parts of another word are to be operated on.
Physical Units	- The units of a variable, such as M/SEC for velocity, DECIBARS for pressure, etc.
Scan	- See data scan, above.
Synchronization	- A synchronization byte is sent by the instrument to establish scan boundaries. An error condition is recorded if the synchronization byte is not detected where expected.
TCI	- Two's complement integer.
Variable	- A piece of information contained in one HP word in a data scan (q.v), and having a variable descriptor (q.v) in a Scale Factor record. Oceanographic data is stored as variables, to be scaled by the scaling factors in the variable descriptors.
Variable Descriptor	- A collection of fields in a Scale Factor record, containing descriptive and scaling factor information for a particular variable.
Word	- See HP word, above.

HP Data Types

ASCII

American Standard Code for Information Interchange is an eight-level code with seven bits plus parity. ASCII fields are used for comments, data labels, and mnemonic identifiers (variable type, ship name). The first two are free field, but a fixed number of characters, and padded with blanks. The latter are fixed at two characters (an HP word) and can be used by programs for testing and searching.

Two's Complement Integer

This is a standard HP 16-bit two's complement integer. Its range is -32768 to 32767, and the default value, used if a field is unfilled for some reason, is zero (all bits set to 0).

Unsigned Integer

This 16-bit integer format is used for all the actual data in the two types of data records. Its range is from zero (all bits set to 0) to 65535 (all bits set to 1). For values from 0 to 32767 the USI and TCI representations are identical. For numbers greater than 32767 the sign of the TCI is interpreted as the most significant bit in the conversion to floating point. See Appendix I, HP data manipulation function R.

Floating Point (Hewlett-Packard 2100 Series 32-bit FP)

These fields are 32 bits (two HP words) with a 23-bit fraction, a seven-bit exponent, and sign bits for both fraction and exponent. Their approximate range is $\pm 10^{38}$ to $\pm 10^{-38}$, and they are used for scaling the data words. See Appendix I, data conversion to physical units, for example.

Appendix III

Block Diagrams of Record Layouts

APPENDIX III

BLOCK DIAGRAMS OF RECORD LAYOUTS

I. TAPE HEADER

1	KEYWORD = 0	PROJECT CODE	YEAR OF EDIT	MONTH OF EDIT	DAY OF EDIT	TAPE NAME (1)	6
7	TAPE NAME (2)	SOURCE TAPE NAME		FORMAT VERSION	UNUSED		
	//						//
55	COMMENTS: TAPE OBJECTIVE						90

II. STATION FILE HEADER

1	KEYWORD = -2 or -3	PROJECT CODE	SHIP CODE	CRUISE NUMBER	STATION NUMBER	DATA VERSION FLAG	6
7	YEAR OF STATION	MONTH OF STATION (START)	DAY OF STATION (START)	TIME (START)	LATITUDE (START) DEGREES	MINUTES * 100	12
13	LONGITUDE (START) DEGREES	MINUTES *100	WORDS PER CTD DATA SCAN	100 * NUMBER SCANS PER SECOND	TIMER PULSE FREQUENCY	PRESSURE SAMPLING INTERVAL	18
19	LATITUDE (END) DEGREES	MINUTES *100	LONGITUDE (END) DEGREES	MINUTES * 100	TIME (END)	MINIMUM PRESSURE	24
25	MAXIMUM PRESSURE	JULIAN DAY NUMBER		INSTRUMENT NUMBER	QUALITY FLAG	YEAR OF EDIT	30
31	MONTH OF EDIT	DAY OF EDIT	NUMBER OF WATER SAMPLES	METHOD OF DETERMINING POSITION	WIND SPEED	WATER DEPTH	36
37	STATION TYPE	CAST NUMBER					
55	NAME OF PERSON CREATING FILE						
	COMMENTS PERTAINING TO FILE						
	PROGRAM VERSION						90

III. SCALE FACTOR RECORD

PROGRAM CONTROL WORDS:

1	KEYWORD =4, 5, or -6	NUMBER OF VARIABLES IN RECORD	LENGTH OF DESCRIPTOR	DATA SCAN LENGTH	NO. OF F.P. IN DESCRIPTOR	UNUSED	6
7	UNUSED						
	VARIABLE DESCRIPTOR 1: APPLIES TO FIRST H.P. WORD OF DATA SCAN						
	VARIABLE DESCRIPTOR 2: APPLIES TO SECOND H.P. WORD OF DATA SCAN						

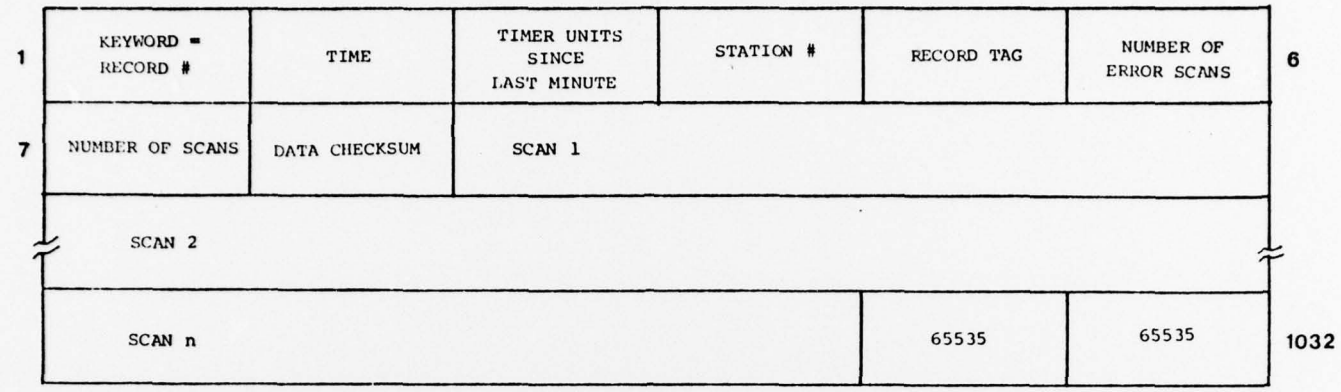
VARIABLE DESCRIPTOR N:
APPLIES TO Nth H.P.
WORD OF DATA SCAN

VARIABLE NAME (ENGLISH)				DIMENSIONAL UNITS OF VARIABLE	
DIMENSIONAL UNITS OF VARIABLE (continued)			VARIABLE IDENTIFIER	LAG CORRECTION WINDOW LENGTH	QUALITY FLAG
NUMBER OF BITS IN VARIABLE	DELTA EDIT CRITERIA	SENSOR NUMBER	YEAR CALIBRATION	MONTH CALIBRATION	DAY CALIBRATION
LOCATING SIGN BIT WORD	SIGN BIT MASK	LEAST SIGNIFICANT BIT LOCATION	LEAST SIGNIFICANT BIT MASK	DIGITIZING RATE	DATA MASK
≈					≈
ATTRIBUTE II	ATTRIBUTE I	SLOPE	BIAS	LAG OF SENSOR	1032

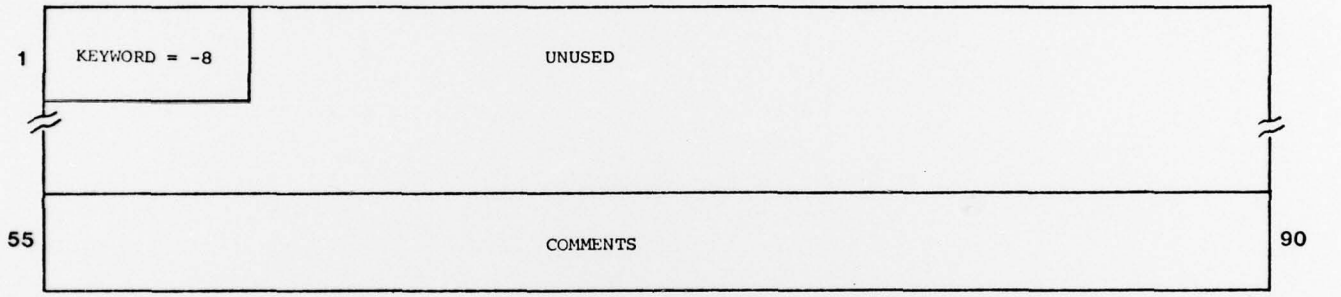
IV. WATER SAMPLE DATA RECORD

1	KEYWORD = -7	UNUSED		6
7	NUMBER OF SCANS	DATA CKECKSUM	SCAN 1	
	SCAN 2			≈
	SCAN n	65535	65535	1032

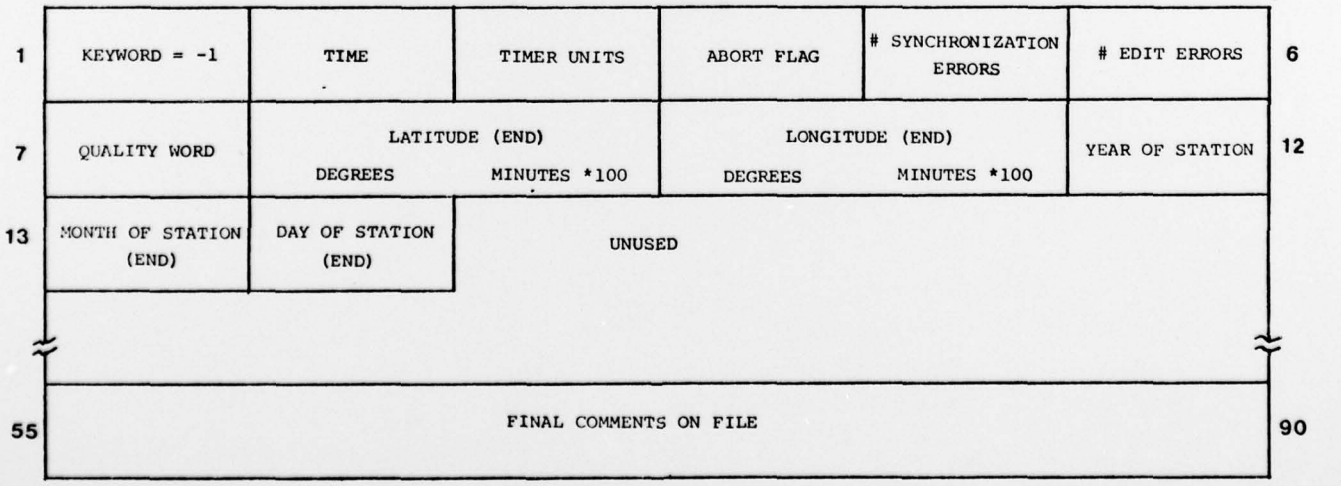
V. CTD DATA RECORD



VI. COMMENT RECORD



VII. FILE TRAILER RECORD



Acknowledgements

We wish to thank the following people for their comments on the report and ideas on the data format: John Dunlap, Nick Fofonoff, Dan Georgi, Terry Joyce, Mary Hunt, John Maltais, Robert McDevitt, George Power, and Jacqueline Webster. This work was supported by ONR Contract N00014-76-C-0197, NR083-400.

References

1. Brown, Neil, 1974. A precision CTD microprofiler. Proceedings, IEEE Ocean 74 Conference, 2, 270-278.

2. Brown, N. L. and G. K. Morrison, 1978. WHOI/Brown Conductivity, Temperature and Depth Microprofiler. WHOI Tech. Rep. 78-23 (unpublished manuscript), 246 p.

3. Global Atmospheric Research Programme, 1974, The International Data Management Plan (Part I) for the GARP Tropical Experiment prepared by T. C. de la Moriniere. No 13, 224 p.

4. Hewlett-Packard, 1972. A Pocket Guide to the 2100 Computer, Hewlett-Packard Co., Cupertino, California.

5. Maltais, John, 1969. A nine-channel digital magnetic tape format for storing oceanographic parameters. WHOI Technical Rep. 69-55 (unpublished manuscript), 11 p.

6. Tantzen, Robert G., 1963. Conversions between calendar date and Julian Day Number, Communications of the ACM, 6(8)444.

7. Tollios, C. D., G. H. Power, and D. J. Ekstrand, 1971. Computer Program for Real Time Digital Acquisition of Conductivity, Temperature, and Pressure. WHOI Tech. Memo. 4-71 (unpublished manuscript), 165 p.

MANDATORY DISTRIBUTION LIST

FOR UNCLASSIFIED TECHNICAL REPORTS, REPRINTS, & FINAL REPORTS
PUBLISHED BY OCEANOGRAPHIC CONTRACTORS
OF THE OCEAN SCIENCE AND TECHNOLOGY DIVISION
OF THE OFFICE OF NAVAL RESEARCH

(REVISED FEBRUARY 1978)

1 Director of Defense Research and Engineering
Office of the Secretary of Defense
Washington, D.C. 20301
ATTN: Office Assistant Director (Research)

Office of Naval Research
Arlington, VA 22217
1 ATTN: (Code 460)
1 ATTN: (Code 102-OS)
6 ATTN: (Code 102IP)
1 ATTN: (Code 200)

1 CDR J. C. Harlett, (USN)
ONR Representative
Woods Hole Oceanographic Inst.
Woods Hole, MA 02543

1 Office of Naval Research
Branch Office
495 Summer Street
Boston, MA 02210

Director
Naval Research Laboratory
Washington, D.C. 20375
6 ATTN: Library, Code 2620

1 National Oceanographic Data Center
National Oceanic & Atmospheric Administration
3300 Whitehaven St., N.W.
Washington, D.C. 20235

12 Defense Documentation Center
Cameron Station
Alexandria, VA 22314

Commander
Naval Oceanographic Office
Washington, D.C. 20373
1 ATTN: Code 1640
1 ATTN: Code 70

3 NORDA 430
NSTL Station, MS 39529

2 CO NORDA
NSTL Station, MS 39529

Woods Hole Oceanographic Institution
WHOI-78-43

A DIGITAL TAPE FORMAT FOR WOODS HOLE OCEANOGRAPHIC INSTITUTION CTD DATA by R. C. Millard, A. Blumer, and N. Galbraith. 36 pages. May 1978. Prepared for the Office of Naval Research under Contract N00014-76-C-0197; NR 083-400.

A new digital data tape format has been developed to be used for data from the WHOI/Brown CTD microprofiler. CTD-78 Version 1 as detailed in this report is designed as a flexible and expandable internal data format adapted to the Hewlett-Packard 2100/21MX series 16-bit mini-computers currently used to transcribe and process CTD data at sea. The ten record types presently used in this multifile tape format store CTD and associated water sample data and the labeling information needed to convert the data to physical units. Record types are identified by unique keywords.

1. Magnetic tape format
2. CTD Data Storage
3. 16-bit mini-computer

I. Millard, R. C.
II. Blumer, A.
III. Galbraith, N.
IV. N00014-76-C-0197; NR 083-400

This card is UNCLASSIFIED

Woods Hole Oceanographic Institution
WHOI-78-43

A DIGITAL TAPE FORMAT FOR WOODS HOLE OCEANOGRAPHIC INSTITUTION CTD DATA by R. C. Millard, A. Blumer, and N. Galbraith. 36 pages. May 1978. Prepared for the Office of Naval Research under Contract N00014-76-C-0197; NR 083-400.

A new digital data tape format has been developed to be used for data from the WHOI/Brown CTD microprofiler. CTD-78 Version 1 as detailed in this report is designed as a flexible and expandable internal data format adapted to the Hewlett-Packard 2100/21MX series 16-bit mini-computers currently used to transcribe and process CTD data at sea. The ten record types presently used in this multifile tape format store CTD and associated water sample data and the labeling information needed to convert the data to physical units. Record types are identified by unique keywords.

1. Magnetic tape format
2. CTD Data Storage
3. 16-bit mini-computer

I. Millard, R. C.
II. Blumer, A.
III. Galbraith, N.
IV. N00014-76-C-0197; NR 083-400

This card is UNCLASSIFIED

Woods Hole Oceanographic Institution
WHOI-78-43

A DIGITAL TAPE FORMAT FOR WOODS HOLE OCEANOGRAPHIC INSTITUTION CTD DATA by R. C. Millard, A. Blumer, and N. Galbraith. 36 pages. May 1978. Prepared for the Office of Naval Research under Contract N00014-76-C-0197; NR 083-400.

A new digital data tape format has been developed to be used for data from the WHOI/Brown CTD microprofiler. CTD-78 Version 1 as detailed in this report is designed as a flexible and expandable internal data format adapted to the Hewlett-Packard 2100/21MX series 16-bit mini-computers currently used to transcribe and process CTD data at sea. The ten record types presently used in this multifile tape format store CTD and associated water sample data and the labeling information needed to convert the data to physical units. Record types are identified by unique keywords.

1. Magnetic tape format
2. CTD Data Storage
3. 16-bit mini-computer

I. Millard, R. C.
II. Blumer, A.
III. Galbraith, N.
IV. N00014-76-C-0197; NR 083-400

This card is UNCLASSIFIED

Woods Hole Oceanographic Institution
WHOI-78-43

A DIGITAL TAPE FORMAT FOR WOODS HOLE OCEANOGRAPHIC INSTITUTION CTD DATA by R. C. Millard, A. Blumer, and N. Galbraith. 36 pages. May 1978. Prepared for the Office of Naval Research under Contract N00014-76-C-0197; NR 083-400.

A new digital data tape format has been developed to be used for data from the WHOI/Brown CTD microprofiler. CTD-78 Version 1 as detailed in this report is designed as a flexible and expandable internal data format adapted to the Hewlett-Packard 2100/21MX series 16-bit mini-computers currently used to transcribe and process CTD data at sea. The ten record types presently used in this multifile tape format store CTD and associated water sample data and the labeling information needed to convert the data to physical units. Record types are identified by unique keywords.

1. Magnetic tape format
2. CTD Data Storage
3. 16-bit mini-computer

I. Millard, R. C.
II. Blumer, A.
III. Galbraith, N.
IV. N00014-76-C-0197; NR 083-400

This card is UNCLASSIFIED